

Comparative and Functional Analysis of Cardiovascular-Related Genes



Len Pennacchio, Ph.D.
Genome Sciences Department
Lawrence Berkeley National Laboratory

Sifting Comparative Sequence: Lipid Metabolism Genes and Regulation

Outline

Computational Tools and Databases

- VISTA
- Cardiovascular Gene Resource
 - Examples
- VistaBrowser
 - Human/Mouse Genome Comparison
 - GenomeVISTA

Identification of a Novel Gene (ApoAV)

- Functional Characterization

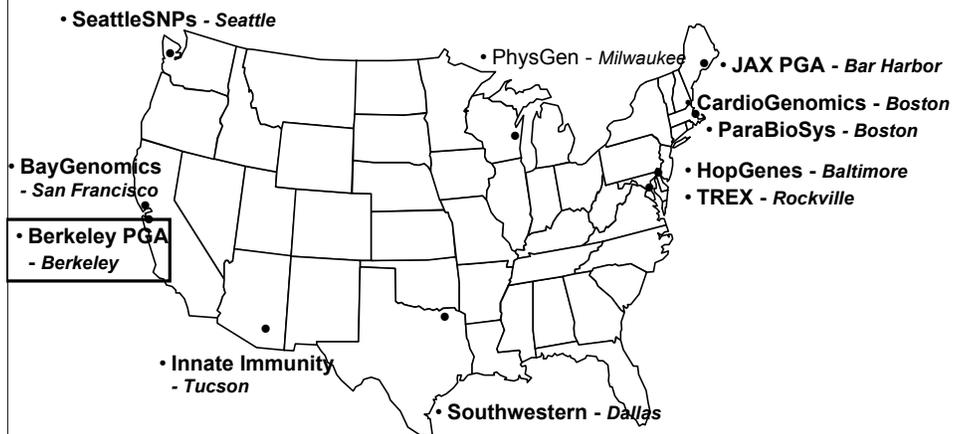
PROGRAMS FOR GENOMIC APPLICATIONS

MISSION: To develop new resources, reagents, and education programs for investigators engaged in NHLBI-related research.

*National Heart, Lung, and Blood
Institute (NHLBI)
National Institutes of Health*



NHLBI PGA Research Network

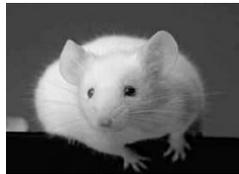


The Human Genome: Feb. 2001



Today: 96% Finished 3% Draft

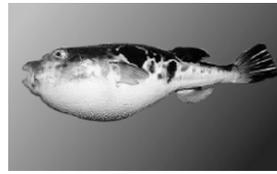
Genomes with Significant Sequence Available



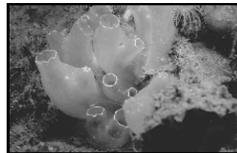
Mouse



Drosophila



Fugu



Ciona (Sea Squirt)



Rat



Zebrafish



Tetraodon

Genomes in the Pipeline



Honeybee



Sea Urchin



Xenopus



Chimpanzee



Chicken



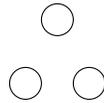
Dog

Why compare the sequence of different organisms?



- Powerful in gene predictions
- Helps in understanding evolution
- Differences between evolutionary closely related species may explain biological differences
- Non-coding sequences conserved between species are reliable guides to regulatory elements

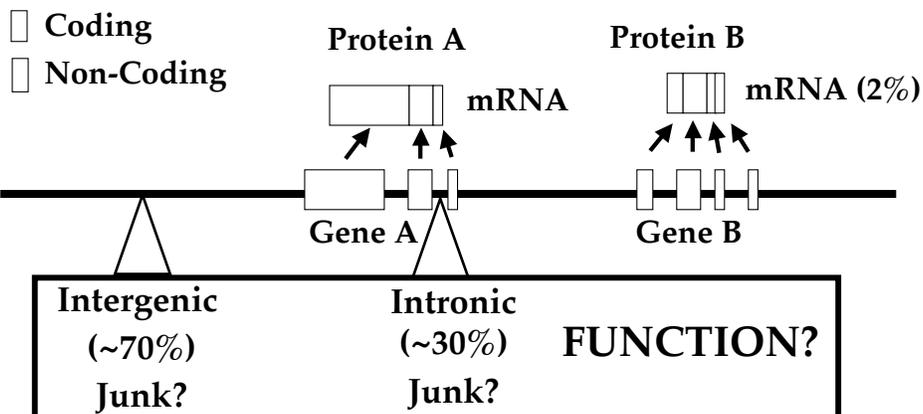
Conserved Sequences are Functionally Important.



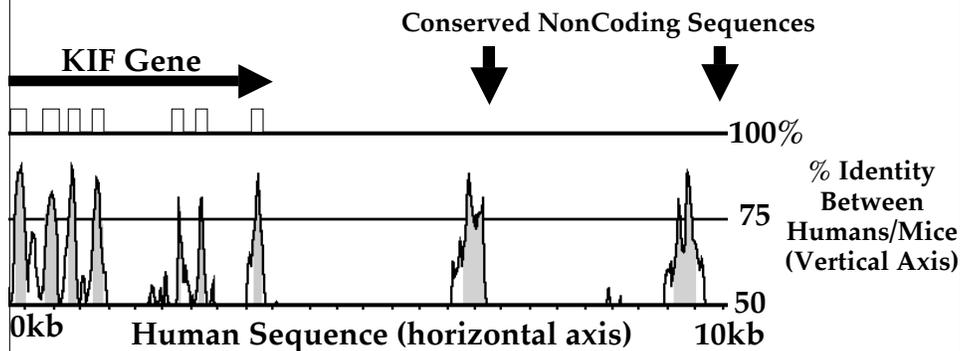
Find & Study these First

Inventory of Mammalian DNA

Coding ~2%
Non-coding ~98%



VISTA Plot (VISual Tool for Alignment)



<http://www-gsd.lbl.gov/vista>

 **PROGRAMS FOR GENOMIC APPLICATIONS** National Heart, Lung, and Blood Institute

[OVERVIEW](#) [TOOLS](#) [DATA](#) [RESEARCHERS](#) [OTHER PGA's](#) [EDUCATION](#)

berkeley PGA

Comparative Genomic Analysis of Cardiovascular Gene Regulation

This project is one of eleven Programs for Genomic Applications (PGAs) funded by the National Heart, Lung, and Blood Institute (NHLBI).

<http://pga.lbl.gov>

Goal:

- Perform comparative sequence analysis for ~250 genes of cardiovascular disease relevance.
 - Examples
- Functionally characterize a subset of conserved elements.
 - Strategy

<http://pga.lbl.gov/>

berkeley ^{PGA}

Cardiovascular Comparative Genomic Database (CVCGD)



This database includes well-studied CV genes, for which an understanding of regulation should provide insights into CV relevant biological issues. While only a fraction of these genes will be characterized in the PGA biological projects over the 4-year time period of this program, the sequence of ~200 genomic intervals containing CV genes will be obtained and comparatively annotated and included in the CVCGD.

The database contains a variety of information for each gene relevant to this project:

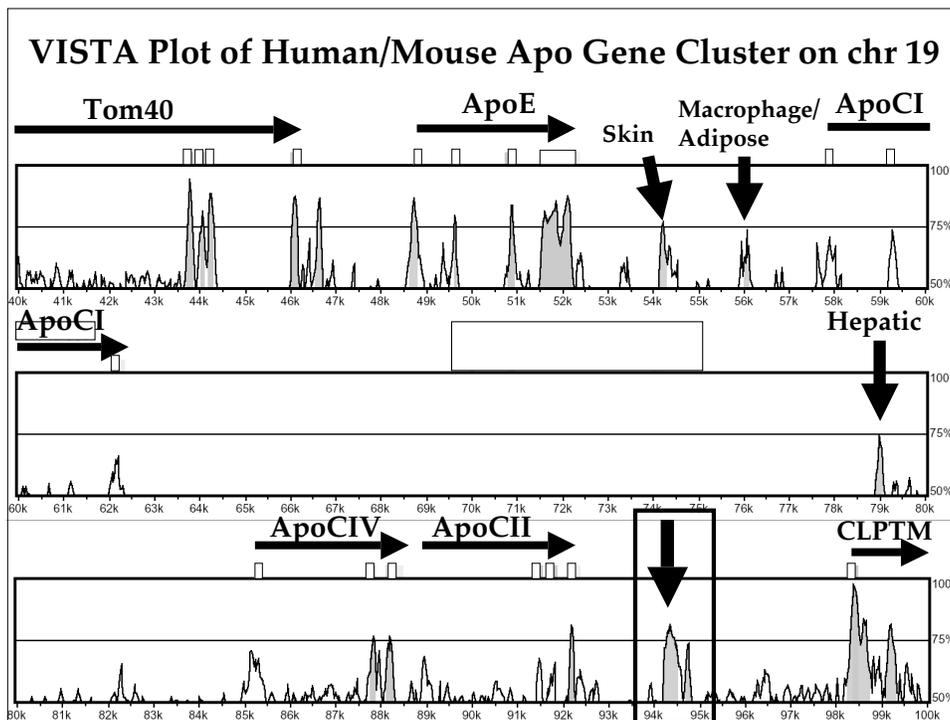
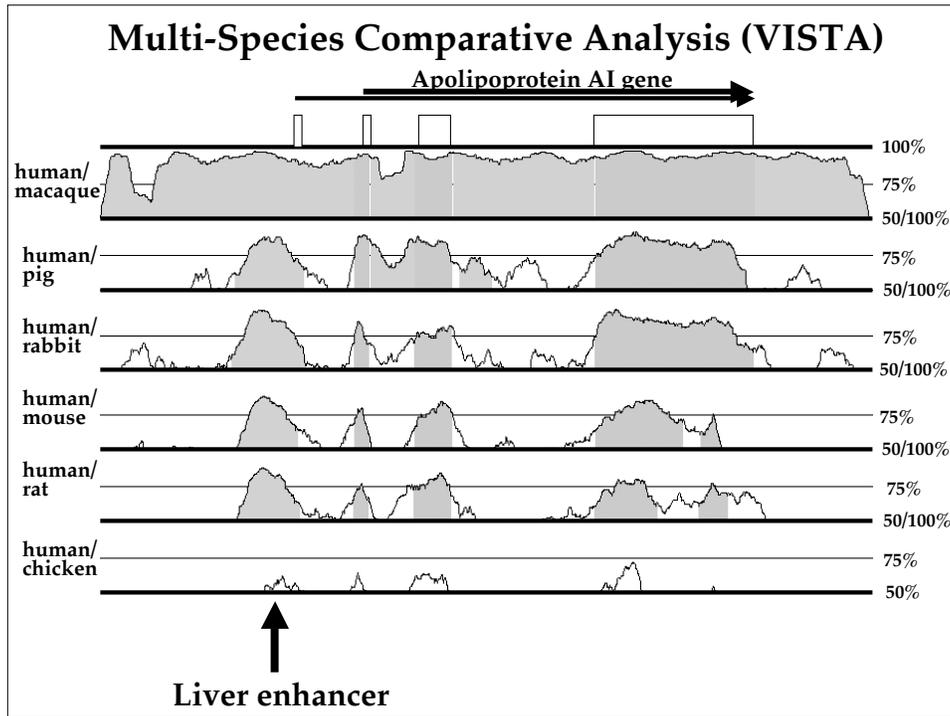
SEARCH the CVCGD

- Gene name;
 - Gene ID in the OMIM database (**OMIM**);
 - Human map location (**HM**);
 - GenBank accession number for human cDNA (**HC**);
 - Mouse map location (**MM**);
 - GenBank accession number for mouse cDNA (**MC**).
- [by gene name and abbreviation](#)
 - [sorted alphabetically](#)
 - [by categories \(groups of diseases\)](#).

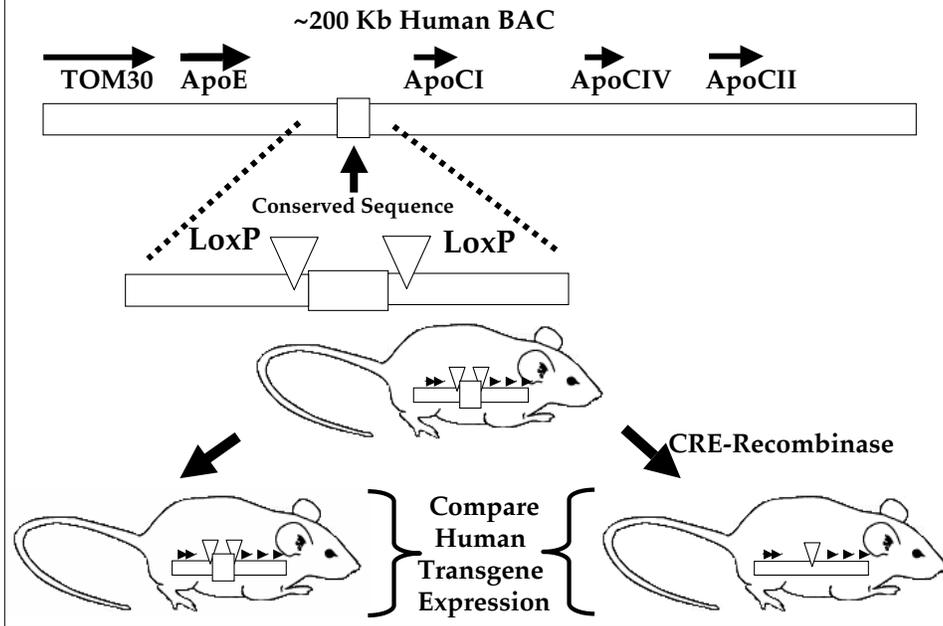
Example

Apolipoprotein A1 (APOA1)

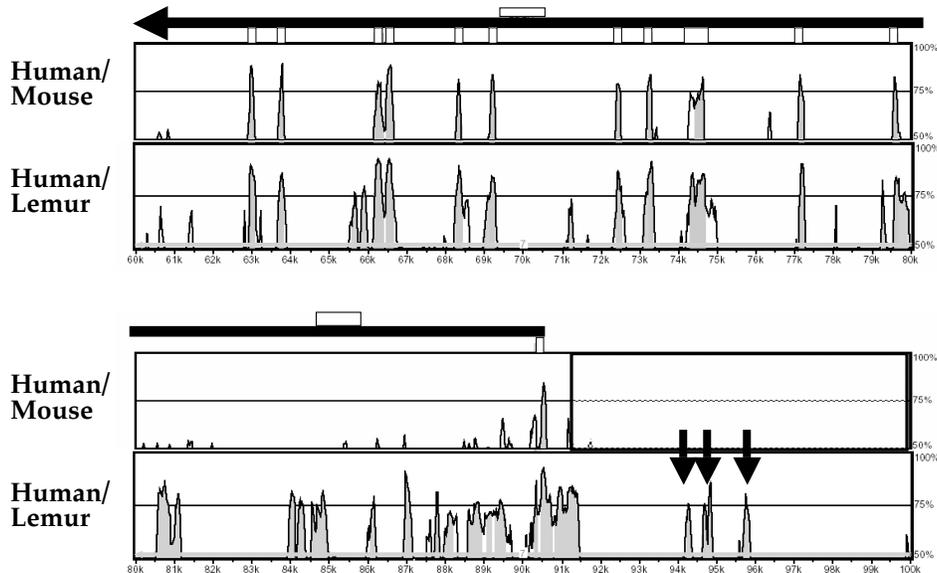
- Category: Atherosclerosis
- Gene ID in the OMIM database: [107680](#)
- Human map location: 11q23
- GenBank accession number for human cDNA: [NM_000039](#)
- Mouse map location: 9
- GenBank accession number for mouse cDNA: [NM_009692](#)
- [Annotation of the human sequence \(Explanation of annotation\)](#)
- Human-mouse alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions \(human-mouse\)](#)
- Human-rabbit alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions \(human-rabbit\)](#)
- Human-lemur alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions \(human-lemur\)](#)
- Human-chimp alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions \(human-chimp\)](#)
- Human-baboon alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions \(human-baboon\)](#)
- Human-chicken alignment: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)
- [List of conserved regions \(human-chicken\)](#)
- All alignments displayed on the same scale: [Whole sequence](#) | [1-80000](#) | [80001-160000](#) | [160001-199644](#) | [Printable version \(PDF\)](#)



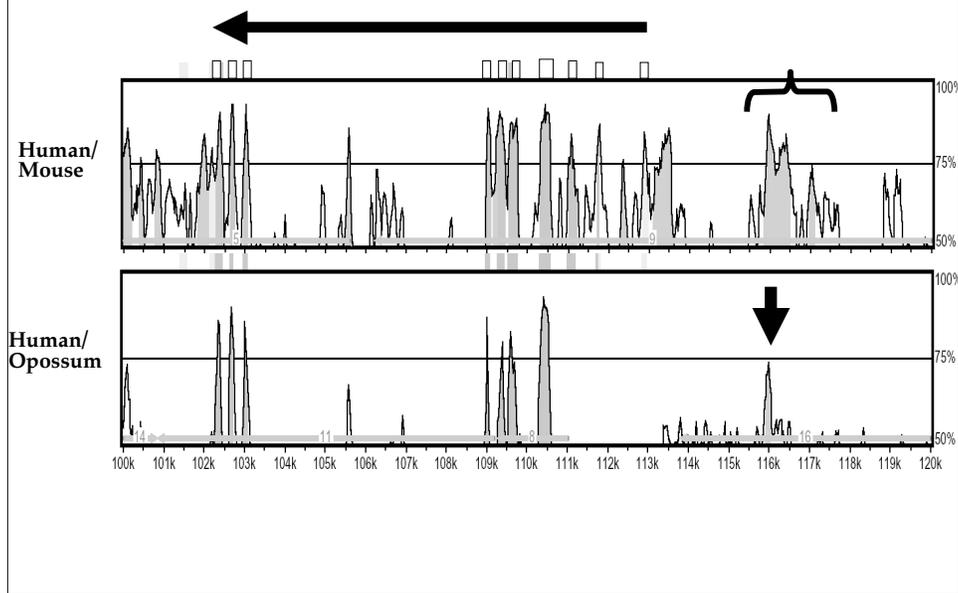
Characterizing Conserved Sequence Function



Low-Density Lipoprotein Receptor (LDLR): (Human/Mouse too distant)



Nuclear Hormone Receptor:LXR-Alpha: (Human/Mouse too close)

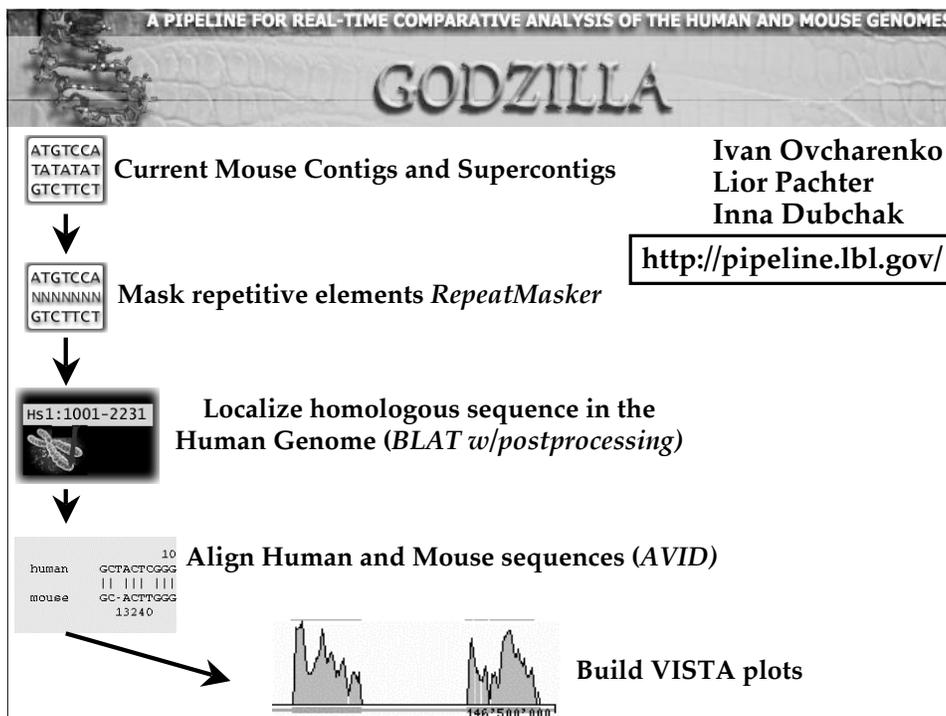
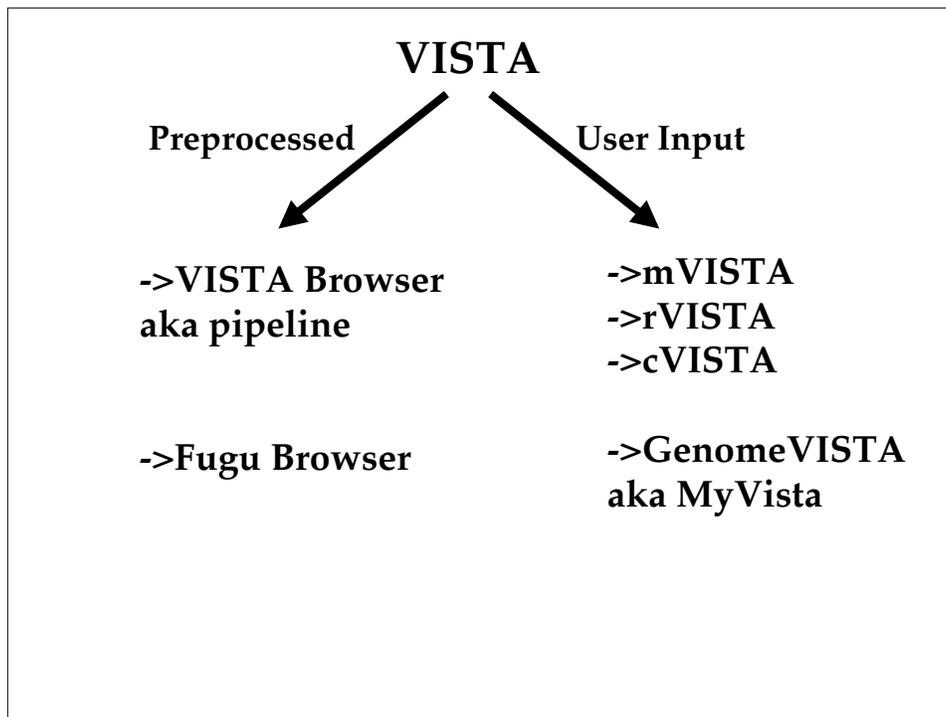


VISTA VISUALIZATION TOOLS FOR ALIGNMENTS
<http://www-gsd.lbl.gov/vista>

mVISTA: main VISTA
-standard comparative sequence plots

rVISTA: regulatory VISTA
-conserved transcription factor binding sites

Availability:
Web-based
Stand Alone Package



VISTA Browser

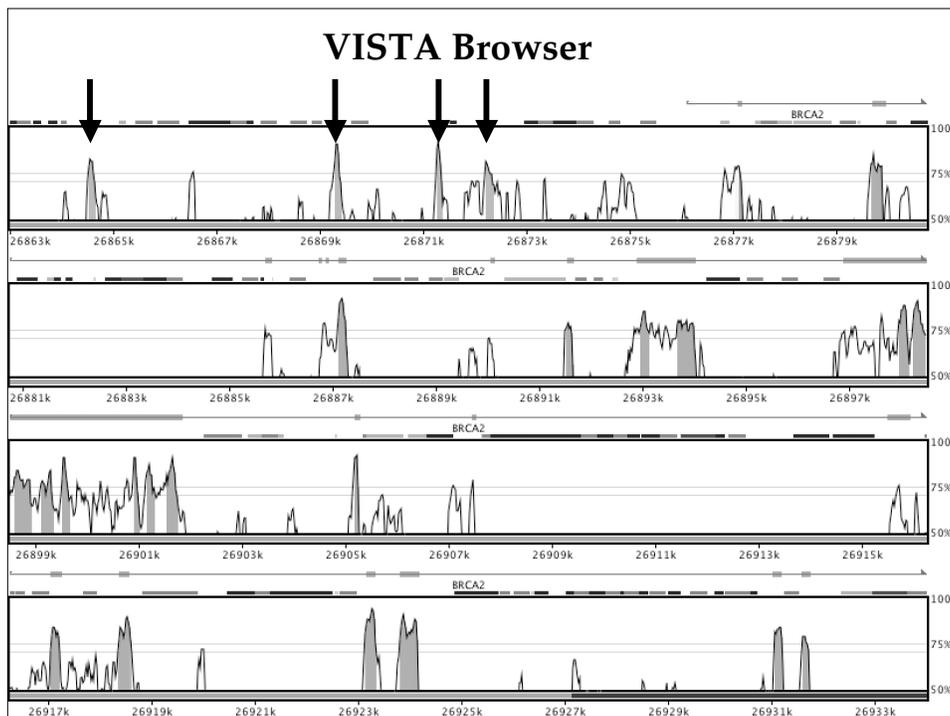
Pre-processed Whole Human/Mouse Genome Comparison

THE BERKELEY GENOME PIPELINE **GODZILLA**

Compare the Human and Mouse Genomes

Please enter a position in the Human Genome (June 2002 draft assembly), select the browser to display comparative analysis results, and press the submit button:

<http://pipeline.lbl.gov/>



GenomeVISTA

Self-Input Sequence Comparison to Human, Mouse or Rat Reference Genomes

Submit a Request

Sequence (choose one of the three options) **Paste a Query Sequence** (FASTA format finished sequences only, 300K max)
Draft sequences can be all entered at once, each config starting with > and the sequence name

Alternatively, you can also select a file or enter a GenBank identification number:

FASTA Or **GenBank**
Text files only. Word documents are **not** accepted. Sequences should be in FASTA format
GenBank Locus: Accession or GI Number



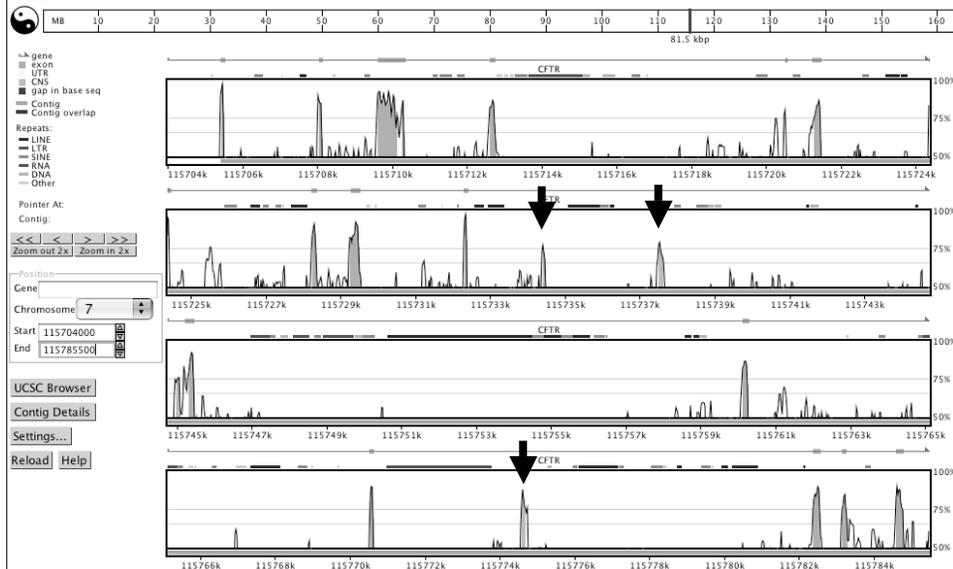
opossum

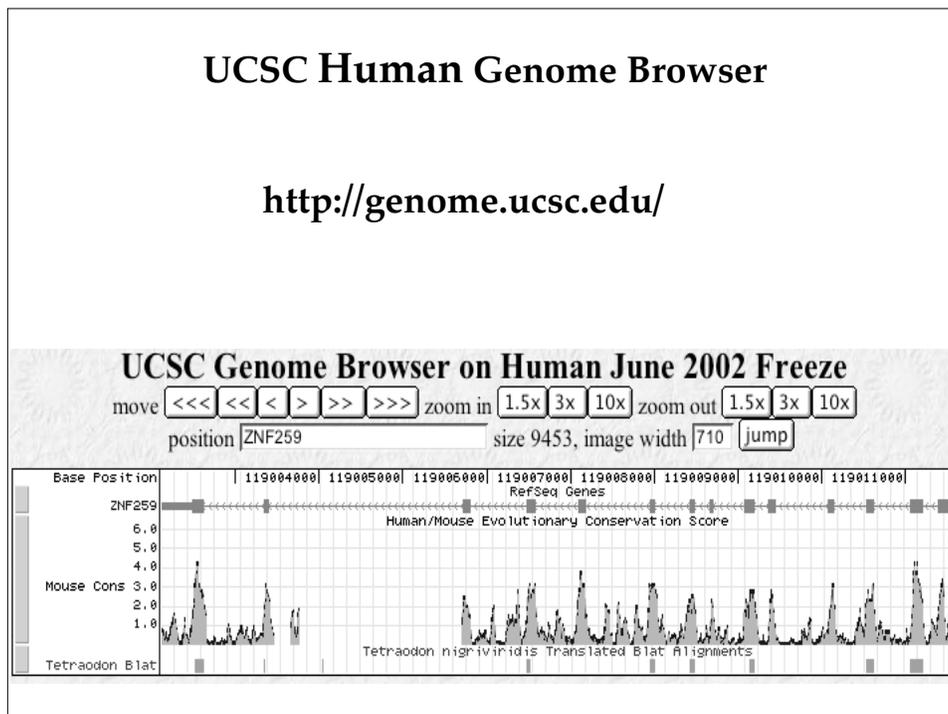
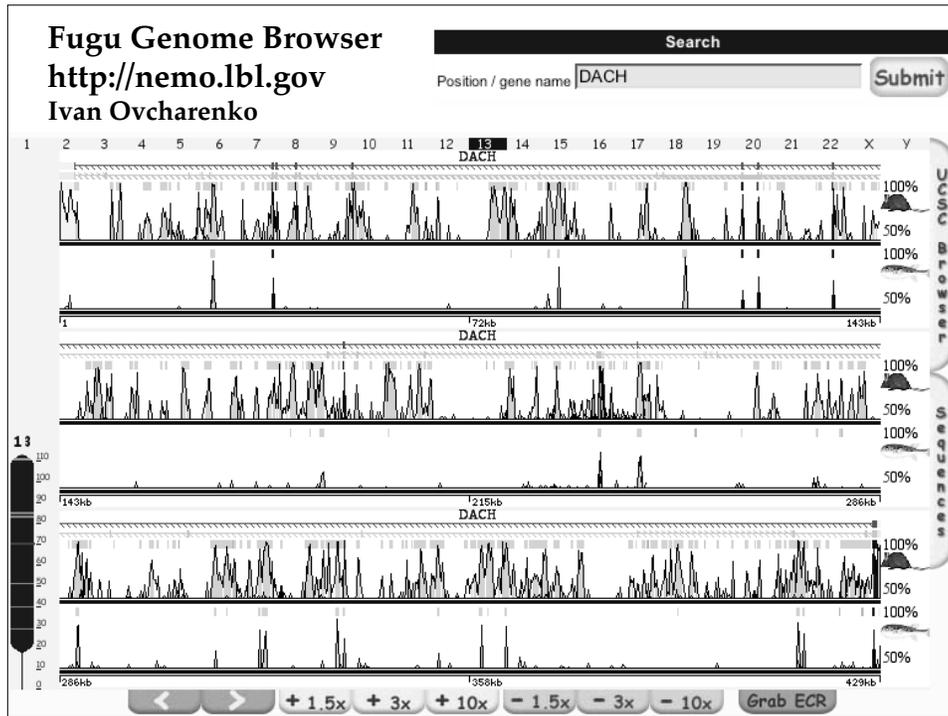
<http://pipeline.lbl.gov/>

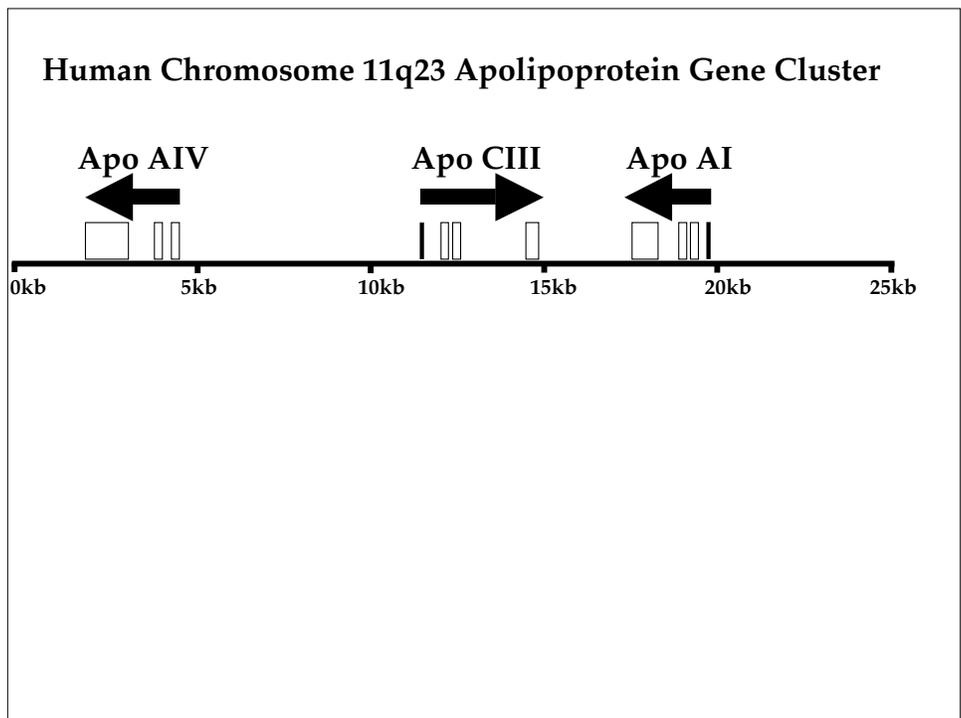
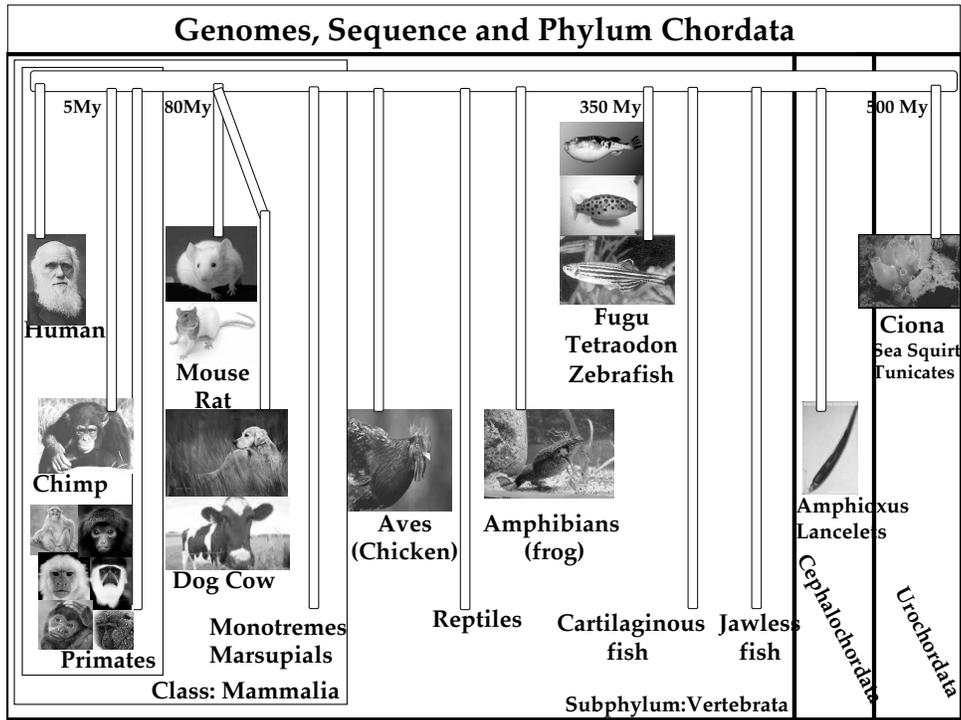


GenomeVISTA

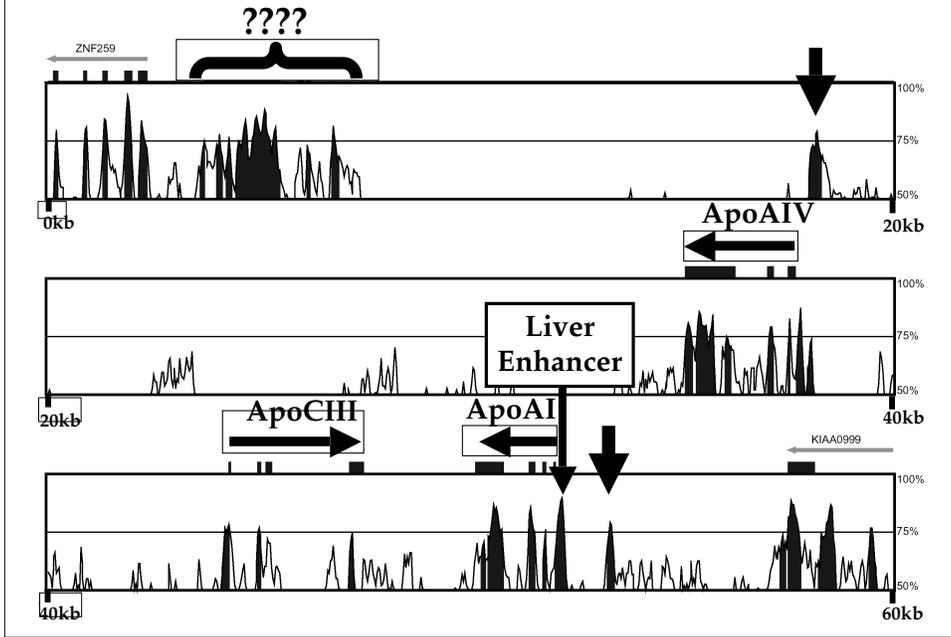
Random Opossum BAC versus Human Genome



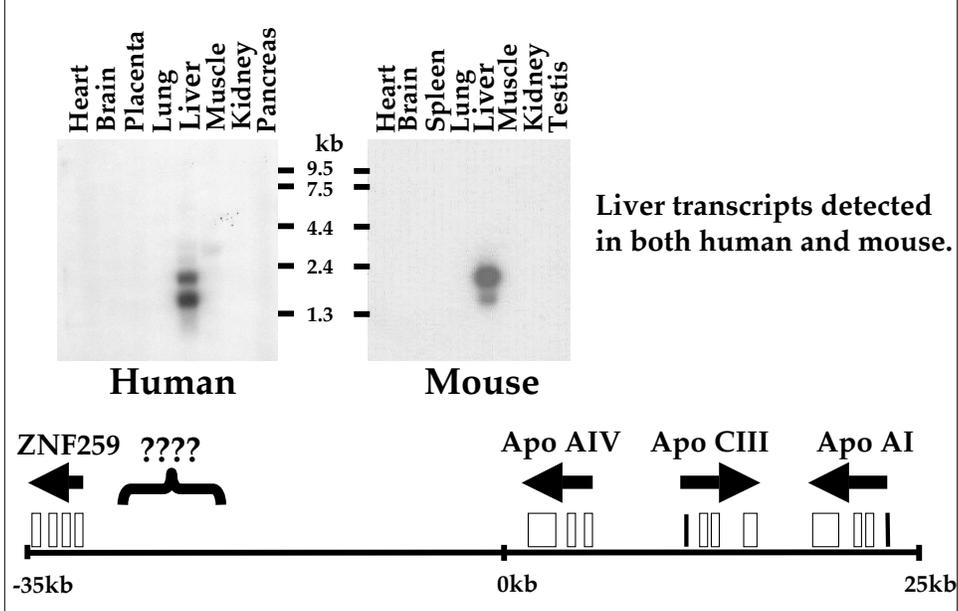




Human/Mouse Apolipoprotein Gene Cluster Sequence Comparison



Northern Blot Analysis of Conserved Sequence



Predicted protein has homology to ApoAIV

predicted protein ---M^AV^LT^MA^LA^LS^S---A^FS^AT^CA^RK^GF^WD^YF^SO^TS^G-D^KG^RV^EO^H
 human apoAIV M^FL^KA^VV^LT^LA^LV^AA^GA^RA^EV^SA^DQ^VA^TV^MW^DY^FS^OL^SN^NA^KE^AV^EH^Q

Q^QK^AR^EP-A^TL^KD^SL^EQ^DL^NN^MK^FL^EK^TR^EL^SG^SE^AP^RI^PO^DP^VG^MR^R
 K^SE^LT^QQ^LN^AL^FO^DK^LG^EV^NT^YA^GD^LQ^RK^LV^FF^AT^EL^HE^RI^AR^DS^EK^LK^E

Q^QE^LE^EV^KA^RL^QE^YM^AE^AH^EL^VG^NL^EG^RO^LK^PY^TM^DL^ME^OV^AL^RV^E
 E^LG^KE^LE^ER^AR^LL^EH^AN^EV^SO^RK^LG^DN^IR^EL^QO^RL^EP^YA^DQ^LR^TO^VN^TO^A

Q^EL^QE^OR^VV^GE^DT^KA^OL^LG^GV^DE^AW^AL^LO^G---L^QS^RV^HH^TG^RF^KE^L
 E^OR^RO^DP^LA^QR^ME^RV^RE^NA^DS^LQ^AS^TR^PH^AD^EL^KA^KD^QN^VE^EL^LG^R

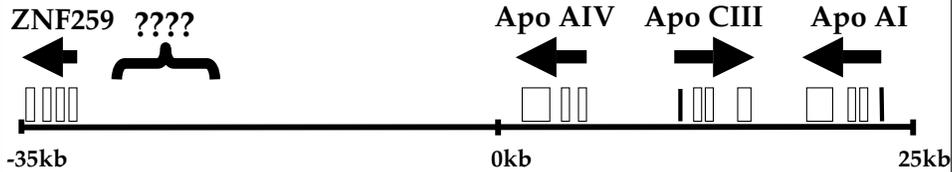
F^HP^YA^ES^LV^SG^TG^RH^VO^EL^HR^SW^AF^HP^AS^PA^NL^SC^QV^LS^RK^LT^LK^AK^L
 L^TP^YA^DE^FK^VK^LD^OT^VE^LR^RS^LA^EY^AQ^DQ^EL^NQ^EG^LL^FO^KK^NA^E

A^LE^AR^IQ^QN^LD^OL^RE^ET^SR^AF^AG^T----C^TE^GA^GP^DP^QM^LS^EE^VR^OR^L
 E^LA^RI^SA^SA^EL^RO^RL^AP^LA^ED^VR^GN^LK^NT^EG^LQ^KS^LA^EL^GG^HD^OV

Q^AF^FQ^DT^YL^QI^AA^FT^AD^OE^TE^VQ^OO^LA^EP^PP^GH^SF^AF^EF^QQ^TD^SG^K
 E^EF^RR^RV^EP^YG^EN^EN^AL^VQ^MQ^OL^RO^KL^EH^AG^DV^EL^SF^LE^KD^LR^DK

V^LS^KL^QA^RL^DL^WE^DI^TH^SL^HQ^GH^SL^GD^P-----
 V^NS^FF^ST^FK^KE^SO^KL^SL^EL^EQ^QE^QQ^QE^QQ^QE^QV^QM^LA^PL^ES

Identity: 26%
 Similarity: 45%



Predicted protein has homology to ApoAIV

predicted protein ---M^AV^LT^MA^LA^LS^S---A^FS^AT^CA^RK^GF^WD^YF^SO^TS^G-D^KG^RV^EO^H
 human apoAIV M^FL^KA^VV^LT^LA^LV^AA^GA^RA^EV^SA^DQ^VA^TV^MW^DY^FS^OL^SN^NA^KE^AV^EH^Q

Q^QK^AR^EP-A^TL^KD^SL^EQ^DL^NN^MK^FL^EK^TR^EL^SG^SE^AP^RI^PO^DP^VG^MR^R
 K^SE^LT^QQ^LN^AL^FO^DK^LG^EV^NT^YA^GD^LQ^RK^LV^FF^AT^EL^HE^RI^AR^DS^EK^LK^E

Q^QE^LE^EV^KA^RL^QE^YM^AE^AH^EL^VG^NL^EG^RO^LK^PY^TM^DL^ME^OV^AL^RV^E
 E^LG^KE^LE^ER^AR^LL^EH^AN^EV^SO^RK^LG^DN^IR^EL^QO^RL^EP^YA^DQ^LR^TO^VN^TO^A

Q^EL^QE^OR^VV^GE^DT^KA^OL^LG^GV^DE^AW^AL^LO^G---L^QS^RV^HH^TG^RF^KE^L
 E^OR^RO^DP^LA^QR^ME^RV^RE^NA^DS^LQ^AS^TR^PH^AD^EL^KA^KD^QN^VE^EL^LG^R

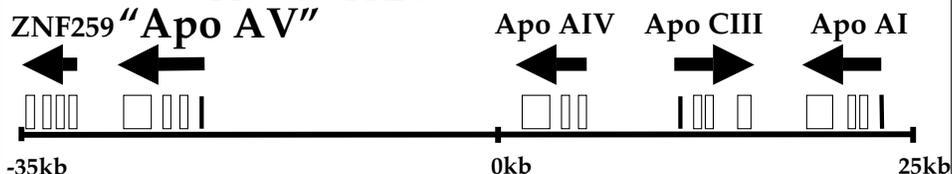
F^HP^YA^ES^LV^SG^TG^RH^VO^EL^HR^SW^AF^HP^AS^PA^NL^SC^QV^LS^RK^LT^LK^AK^L
 L^TP^YA^DE^FK^VK^LD^OT^VE^LR^RS^LA^EY^AQ^DQ^EL^NQ^EG^LL^FO^KK^NA^E

A^LE^AR^IQ^QN^LD^OL^RE^ET^SR^AF^AG^T----C^TE^GA^GP^DP^QM^LS^EE^VR^OR^L
 E^LA^RI^SA^SA^EL^RO^RL^AP^LA^ED^VR^GN^LK^NT^EG^LQ^KS^LA^EL^GG^HD^OV

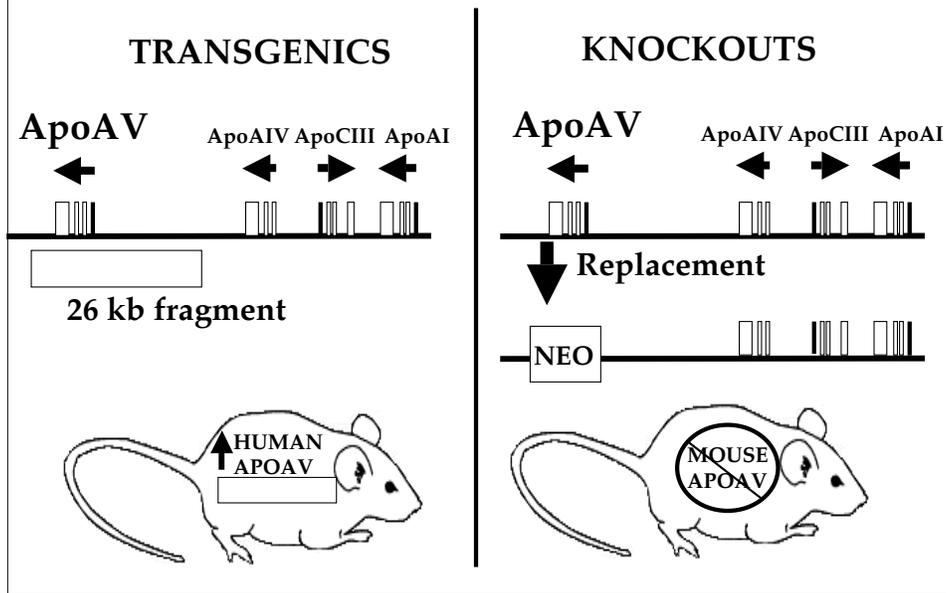
Q^AF^FQ^DT^YL^QI^AA^FT^AD^OE^TE^VQ^OO^LA^EP^PP^GH^SF^AF^EF^QQ^TD^SG^K
 E^EF^RR^RV^EP^YG^EN^EN^AL^VQ^MQ^OL^RO^KL^EH^AG^DV^EL^SF^LE^KD^LR^DK

V^LS^KL^QA^RL^DL^WE^DI^TH^SL^HQ^GH^SL^GD^P-----
 V^NS^FF^ST^FK^KE^SO^KL^SL^EL^EQ^QE^QQ^QE^QQ^QE^QV^QM^LA^PL^ES

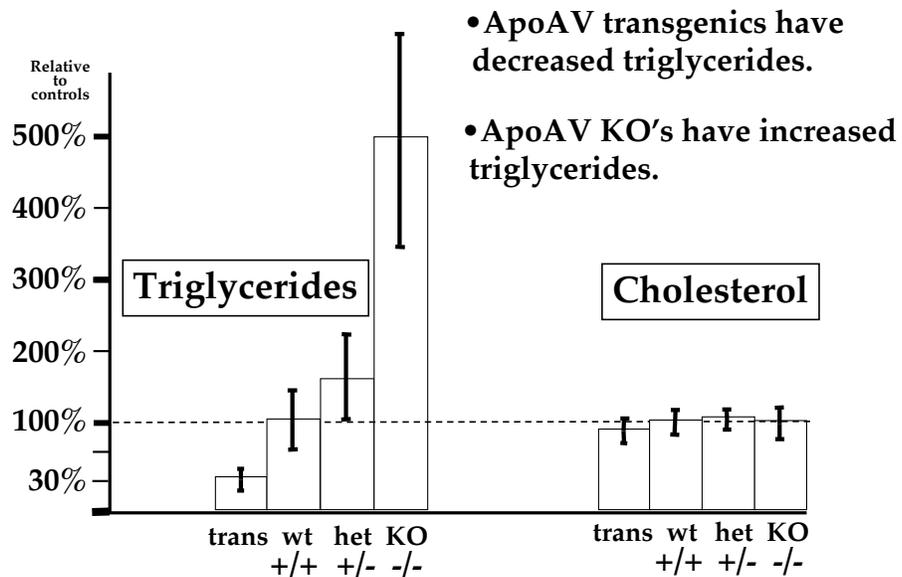
Identity: 26%
 Similarity: 45%



ApoAV Mouse Studies

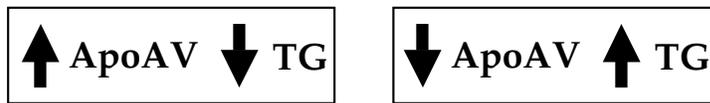


ApoAV Transgenic and Knockout Plasma Levels

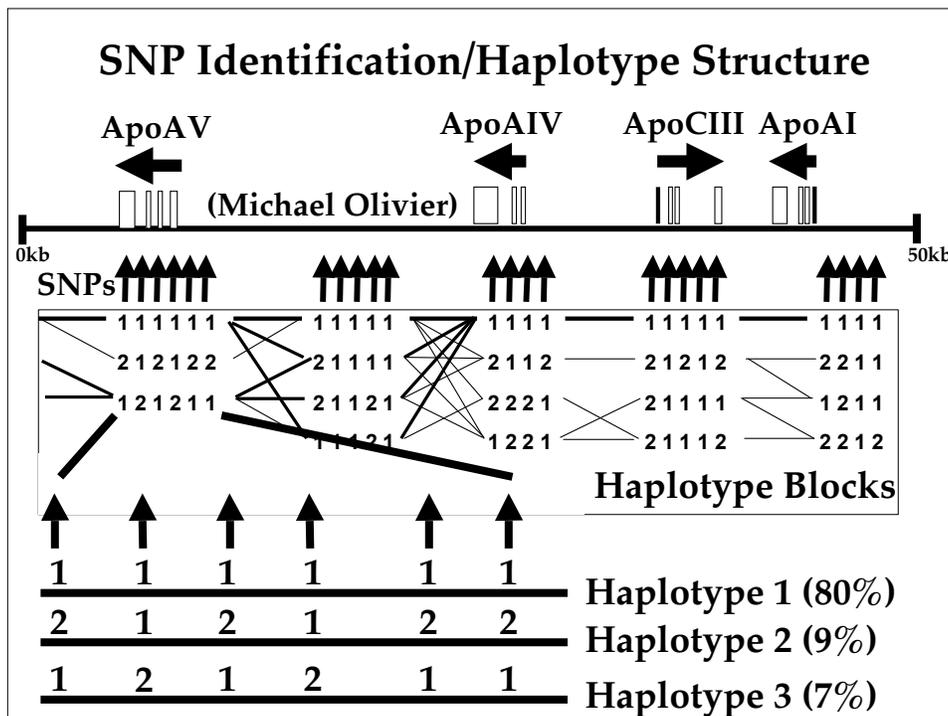


Summary I: ApoAV

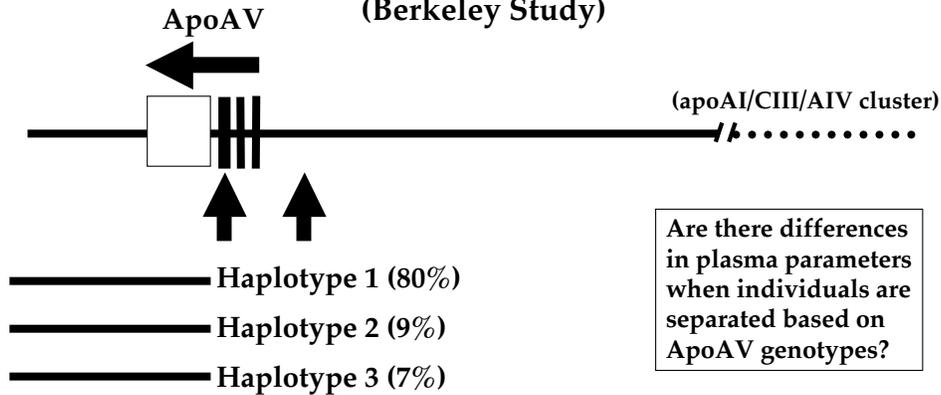
- A new apolipoprotein belonging to the ApoAI/CIII/AIV gene cluster.
- Expressed in the liver & associates with HDL/VLDL.
- An important modulator of triglycerides (TG) in mice.



Is ApoAV involved in human biology/disease?



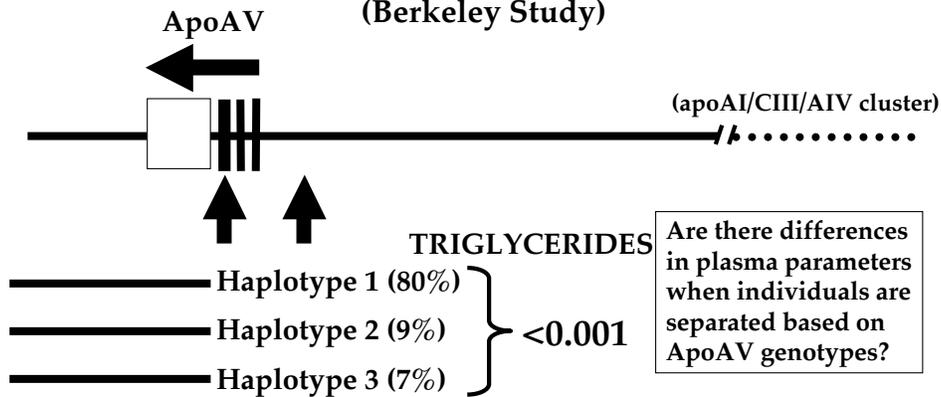
Association study I: ApoAV polymorphisms and plasma parameters (Berkeley Study)



Are there differences in plasma parameters when individuals are separated based on ApoAV genotypes?

- Genotyped 500 normal individuals phenotyped for plasma:
- Triglycerides
 - IDL, LDL, HDL, VLDL Mass
 - HDL, LDL Cholesterol
 - ApoAI, ApoB

Association study I: ApoAV polymorphisms and plasma parameters (Berkeley Study)



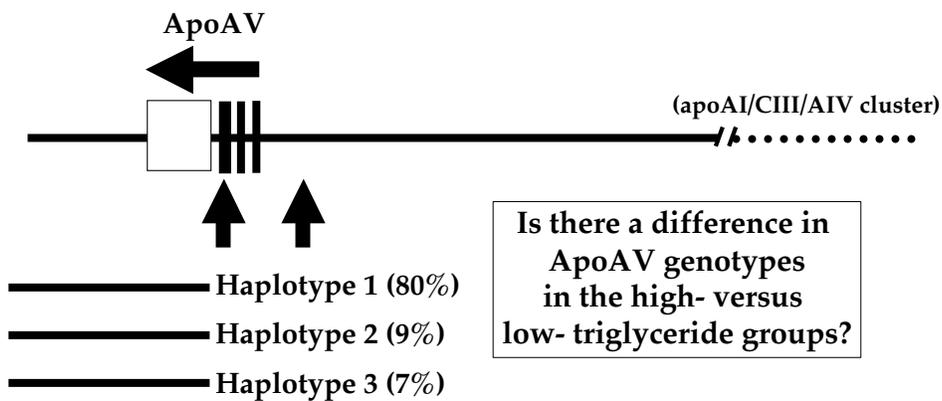
- Genotyped 500 normal individuals phenotyped for plasma:
- TRIGLYCERIDES*
 - IDL, LDL, HDL, VLDL Mass
 - HDL, LDL Cholesterol
 - ApoAI, ApoB*

Association Studies



Is this finding reproducible????

Association study II: ApoAV polymorphisms and plasma parameters

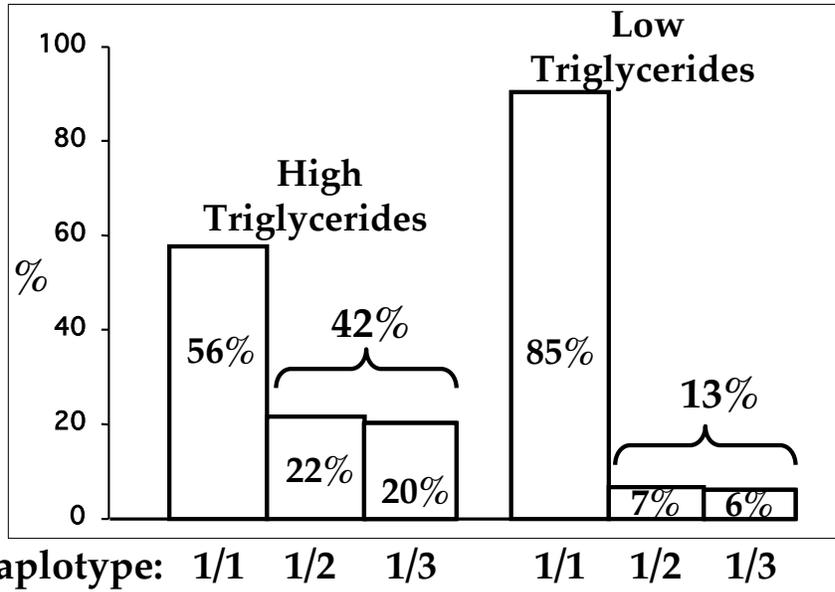


Genotyped 460 individuals stratified based on triglyceride levels.

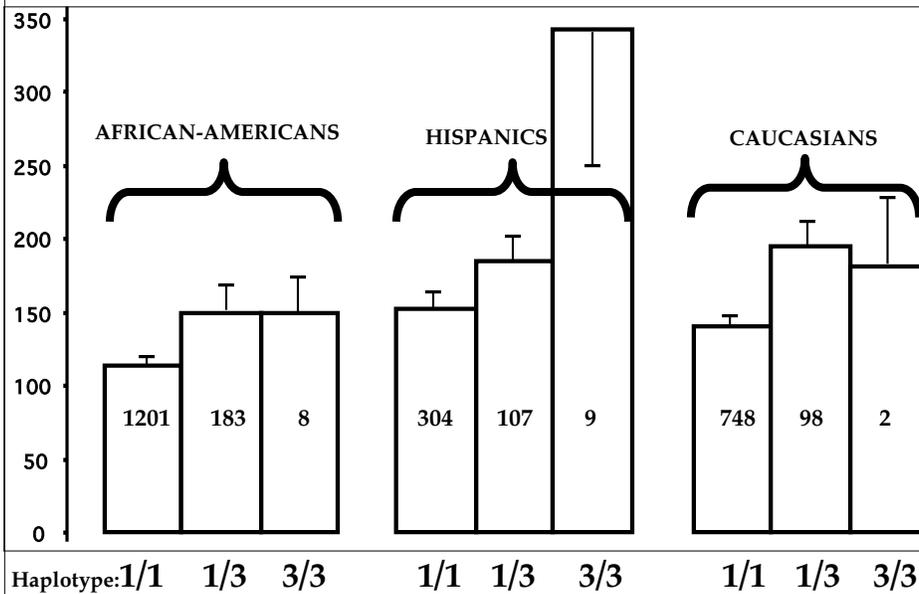
High Triglycerides: Avg 340 mg/dl

Low Triglycerides: Avg 50 mg/dl

**Association study II: ApoAV polymorphisms and plasma parameters
(Dallas Study)**



Association Study III: Dallas, TX

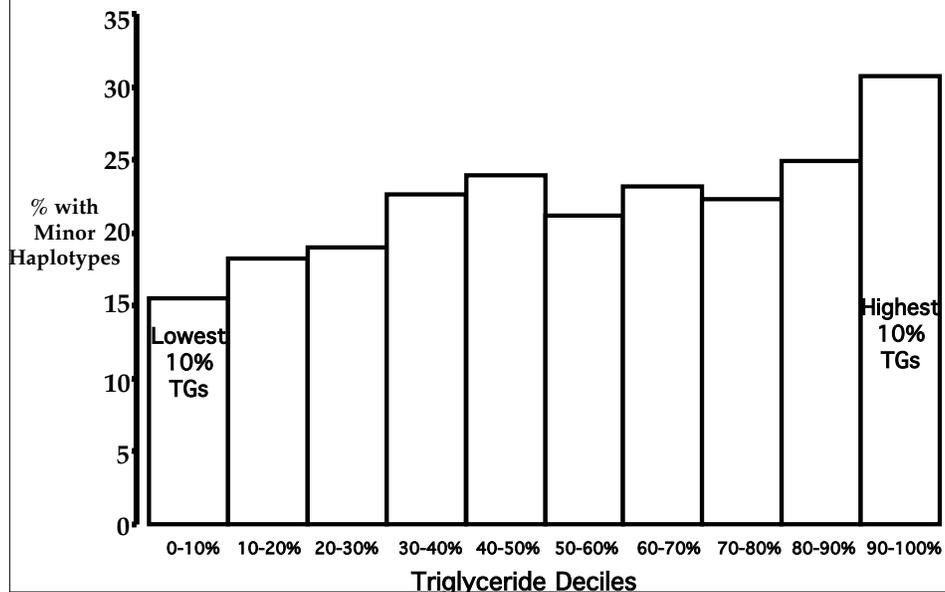


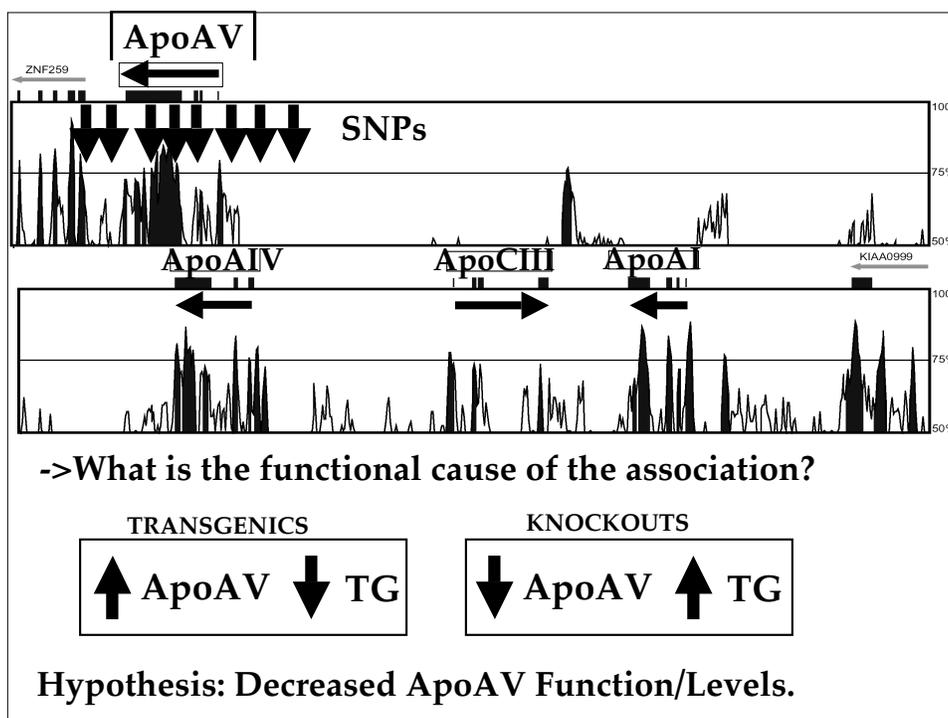
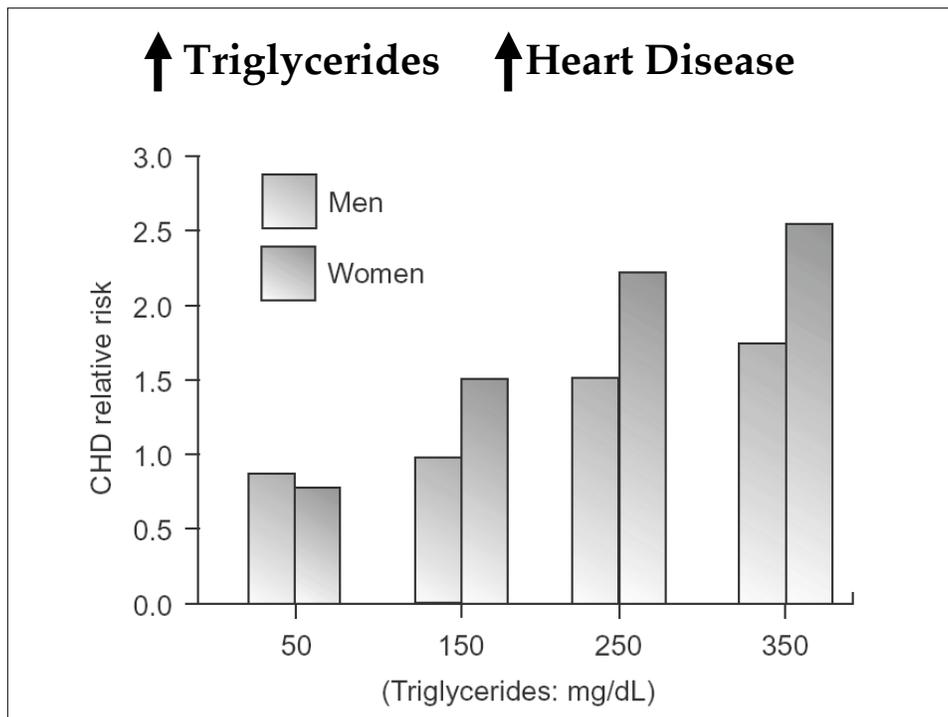
ApoA5 and Triglyceride Levels

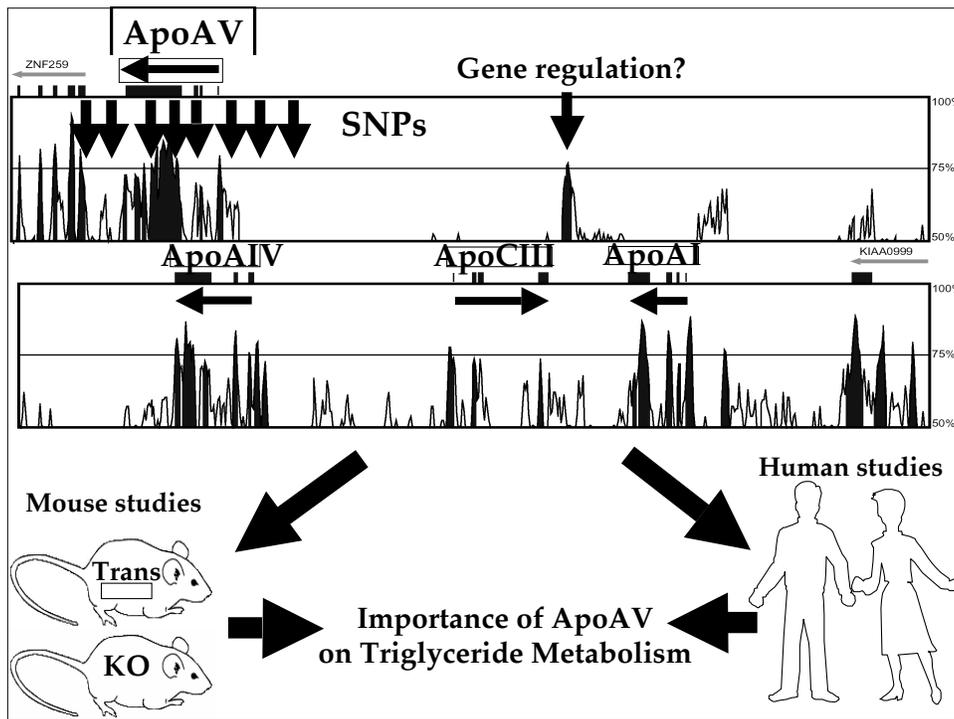
An example of common human variation contributing to a quantitative phenotype

Ethnicity:	Carriers of Minor Haplotype 2 and/or 3:
Caucasian	24%
African American	36%
Hispanic	51%

Common human variation contributing to a quantitative phenotype:
Analysis of 3000 Caucasian individuals separated by triglycerides







Acknowledgements

LBNL

Edward Rubin
Nadine Barouhk
Elaine Gong
Jennifer Akiyama
Kathryn Houston
Keith Lewis
Willow Dean
Jan-Fang Cheng
Inna Dubchak
Lior Pachter
Ivan Ovcharenko
Jody Schwartz
Veena Afzal
Xinli Yang
Ronald Krauss
Patricia Blanche
Laura Holl
Joseph Orr

UT-SW

Jonathan Cohen
Helen Hobbs
Jaroslav Hubacek

Rayne Institute

Philippa Talmud
Steve Humphries

Pasteur Institute-Lille

Jamila Fruchart
Jean-Charles Fruchart

MCW

Michael Olivier

NIH/NHLBI

<http://pga.lbl.gov>

UCSF:

Brian Black

<http://www-gsd.lbl.gov/>